

Best Available Copy



Europäisches
Patentamt

European
Patent Office

Office européen
des brevets

REC'D 04 MAY 2004

WIPO

PCT

NL 0304129
PCT/IB04/50436
II

Bescheinigung

Certificate

Attestation

Die angehefteten Unterla-
gen stimmen mit der
ursprünglich eingereichten
Fassung der auf dem näch-
sten Blatt bezeichneten
europäischen Patentanmel-
dung überein.

The attached documents
are exact copies of the
European patent application
described on the following
page, as originally filed.

Les documents fixés à
cette attestation sont
conformes à la version
initialement déposée de
la demande de brevet
européen spécifiée à la
page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

03076166.2 ✓

**PRIORITY
DOCUMENT**

SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk



Anmeldung Nr:
Application no.: 03076166.2 ✓
Demande no:

Anmeldetag:
Date of filing: 18.04.03 ✓
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

Koninklijke Philips Electronics N.V.
Groenewoudseweg 1
5621 BA Eindhoven
PAYS-BAS

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se référer à la description.)

Audio signal synthesizing

In Anspruch genommene Priorität(en) / Priority(ies) claimed / Priorité(s)
revendiquée(s)
Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

G10L/

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of
filing/Etats contractants désignées lors du dépôt:

AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IT LU MC NL
PT RO SE SI SK TR LI

Audio signal synthesizing

The invention relates to synthesizing an audio signal, and in particular to an apparatus supplying an output audio signal.

5 Erik Schuijers, Werner Oomen, Bert den Brinker and Jeroen Breebaart,
"Advances in Parametric Coding for High-Quality Audio", Preprint 5852, 114th AES
Convention, Amsterdam, The Netherlands, 22-25 March 2003 disclose a parametric coding
scheme using an efficient parametric representation for the stereo image. Two input signals
are merged into one mono audio signal. Perceptually relevant spatial cues are explicitly
10 modeled. The merged signal is encoded using a mono parametric encoder. The stereo
parameters Interchannel Intensity Difference (IID), the Interchannel Time Difference (ITD)
and the Interchannel Cross-Correlation (ICC) are quantized, encoded and multiplexed into a
bitstream together with the quantized and encoded mono audio signal. At the decoder side the
bitstream is de-multiplexed to an encoded mono signal and the stereo parameters. The
15 encoded mono audio signal is decoded in order to obtain a decoded mono audio signal m'
(see Fig. 1). From the mono time domain signal, a de-correlated signal is calculated using a
filter D 10 yielding optimum perceptual de-correlation. Both the mono time domain signal m'
and the de-correlated signal d are transformed to the frequency domain. Then the frequency
domain stereo signal is processed with the IID, ITD and ICC parameters by scaling, phase
20 modifications and mixing, respectively, in a parameter processing unit 11 in order to obtain
the decoded stereo pair l' and r' . The resulting frequency domain representations are
transformed back into the time domain.

25 An object of the invention is to advantageously synthesize an output audio
signal on the basis of an input audio signal. To this end, the invention provides a method, a
device, an apparatus and a computer program product as defined in the independent claims.
Advantageous embodiments are defined in the dependent claims.

According to a first aspect of the invention, synthesizing an output audio signal is provided based on an input audio signal, the input audio signal comprising a plurality of input subband signals, wherein at least one input subband signal is transformed from subband domain to frequency domain to obtain at least one respective transformed signal, wherein the at least one input subband signal is delayed and transformed to obtain at least one respective transformed delayed signal, wherein at least two processed signals are derived from the at least one transformed signal and the at least one transformed delayed signal, wherein the processed signals are inverse transformed from frequency domain to subband domain to obtain respective processed subband signals, and wherein the output audio signal is synthesized from the processed subband signals. By providing a subband to frequency transform in a subband, the frequency resolution is increased. Such an increased frequency resolution has the advantage that it becomes possible to achieve high audio quality (the bandwidth of a single sub-band signal is typically much higher than that of critical bands in the human auditory system) in an efficient implementation (because only a few bands have to be transformed). Synthesizing the stereo signal in a subband has the further advantage that it can be easily combined with existing subband based audio coders. Filter banks are commonly used in the context of audio coding. MPEG-1/2 Layer I, II and III all make use of a 32 bands critically sampled sub band filter.

Embodiments of the invention are of particular use in increasing the frequency resolution of the lower sub-bands using Spectral Band Replication ("SBR") techniques.

In an efficient embodiment, a Quadrature Mirror Filter ("QMF") bank is used. Such a filter bank is known per se from Per Ekstrand, "Bandwidth extension of audio signals by spectral band replication", Proc. 1st IEEE Benelux Workshop on Model based Processing and Coding of Audio (MPCA-2002), pp. 53-58, Leuven, Belgium, November 15, 2002. The synthesis QMF filter bank takes the N complex sub band signals as input and generates a real valued PCM output signal. The idea behind SBR is that the higher frequencies can be reconstructed from the lower frequencies using only very little helper information. In practice, this reconstruction is done by means of a complex Quadrature Mirror Filter (QMF) bank. In order to efficiently come to a de-correlated signal in the subband domain, embodiments of the invention use a frequency (or subband index) dependent delay in the subband domain, as disclosed in more detail in the European patent application of the Applicant filed 17APR2003, entitled " Audio signal generation" (Attorney's docket PHNL030447). Because the complex QMF filter bank is not critically sampled no extra provisions need to be taken in order to account for aliasing. Note that in the SBR decoder as

disclosed by Ekstrand, the analysis QMF bank consists of only 32 bands, while the synthesis QMF bank consists of 64 bands, as the core decoder runs at half the sampling frequency compared to the entire audio decoder. In the corresponding encoder however, a 64 bands analysis QMF bank is used to cover the whole frequency range.

5 Fig. 2 shows a block-diagram of a Bandwidth Enhanced (BWE) decoder using the Spectral Band Replication (SBR) technique as disclosed in MPEG-4 standard ISO/IEC 14496-3:2001/14496-3:2001/FDAM1, JTC1/SC29/WG11, Coding of Moving Pictures and Audio, Bandwidth Extension. The core part of the bit stream is decoded using the core decoder, which can e.g. be a standard MPEG-1 Layer III (mp3) or an AAC decoder. Typically such a
10 decoder runs at half the output sampling frequency ($f_s/2$). In order to synchronise the SBR data with the core data a delay 'D' is introduced (288 PCM samples in the MPEG-4 standard). The resulting signal is fed to a 32 bands complex Quadrature Mirror Filter (QMF). This filter outputs 32 complex samples per 32 real input samples and is thus over-sampled by a factor of 2. In the High-Frequency (HF) generator (see Figure 1) the higher frequencies,
15 which are not covered by the core coder, are generated by replicating (certain parts of) the lower frequencies. The output of the high-frequency generator is combined with the lower 32 sub bands into 64 complex sub-band signals. Subsequently the envelope adjuster adjusts the replicated high frequency sub-band signals to the desired envelope and adds additional sinusoidal and noise components as denoted by the SBR part of the bit-stream. The total 64
20 sub-band signals are fed through the 64 bands complex QMF synthesis filter to form the (real) PCM output signal.

Application of additional transforms, in a sub-band channel, introduces a certain delay. In subbands where no transform and inverse transform is included, delays should be introduced to keep alignment of the subband signals. Without special measures, the
25 extra delay in the subband signals so introduced, results in a misalignment (i.e. out of sync) of the core and side or helper data such as SBR data or parametric stereo data. In the case of the sub bands with additional transform/inverse transform and sub bands without additional transform, additional delay should be added to the sub bands without transform. Within SBR, the extra delay caused by the transforming and inverse transforming operation could be
30 deducted from the delay D.

These and other aspects of the invention are apparent from and will be elucidated with reference to the embodiments described hereinafter.

PHNL030459EPS

4

18.04.2003

In the drawings:

Fig. 1 shows a block diagram of a parametric stereo decoder;

Fig. 2 shows a block diagram of an audio decoder using SBR technology;

Fig. 3 shows parametric stereo processing in sub-band domain according to an

5 embodiment of the invention;

Fig. 4 shows a block diagram illustrating the delay caused by transform-inverse transform TT^{-1} of Fig. 3;

Fig. 5 shows an advantageous audio decoder according to an embodiment of the invention, which provides parametric stereo, and

10 Fig. 6 shows an advantageous audio decoder according to an embodiment of the invention, which combines parametric stereo with SBR.

The drawings only show those elements that are necessary to understand the invention.

15

Fig. 3 shows parametric stereo processing in sub-band domain according to an embodiment of the invention. The input signal consists of N input subband signals. In practical embodiments, N is 32 or 64. The lower frequencies are transformed using transform T to obtain a higher frequency resolution, the higher frequencies are delayed using delay D_T to compensate for the delay introduced by the transform. From each sub band signal also a de-correlated sub-band signal is created by means of delay-sequence D_x where x is the sub-band index. The blocks P denote the processing into two sub-bands from one input subband signal, the processing being performed on one transformed version of the input subband signal and one delayed and transformed version of the input subband signal. The processing may comprise mixing, e.g. by matrixing and/or rotating, the transformed version and the transformed and delayed version. The transform T^{-1} denotes the inverse transform. D_T may be split before and after block P. Transforms T may be of different length, typically low frequency has longer transform, this means that additionally a delay should also be introduced in the paths where the transform is shorter than the longest transform. The delay D in front of filter bank may be shifted after filter bank. When it is placed after the filter bank it can be partially removed because the transforms already incorporate a delay. The transform is preferably of the type Modified Discrete Cosine Transform ("MDCT"), although other transforms such as Fast Fourier Transform may also be used. The processing P does usually not give rise to additional delay.

20

25

30

Fig. 4 shows a block diagram illustrating the delay caused by transform-inverse transform TT^{-1} of Fig. 3. In Fig. 4, 18 complex sub-band samples are windowed by a window $h[n]$. The complex signals are then split into the real and imaginary part, which are both transformed using the MDCT into two times 9 real values. The inverse transform of both sets of 9 values again leads to 18 complex sub band samples that are windowed and overlap-added with the previous 18 complex sub-band samples. As illustrated in this figure, the last 9 complex sub-band samples are not fully processed (i.e. overlap-added) leading to an effective delay of half the transform length, i.e. 9 (sub-band) samples. So, the delay in a single sub-band filter should be compensated in all other sub bands where no transformation is applied. However, introducing an extra delay to the subband signals prior to SBR processing (i.e. HF generation and envelope adjustment) results in a misalignment of the core and SBR data. In order to preserve this alignment the PCM delay D as shown in Fig. 2 can be placed just after the M bands complex analysis QMF, which effectively results in a delay of D/M in each subband. Thus, the requirement for alignment of the core and SBR data is that the delay in all subbands amounts to D/M . Therefore as long as the delay DT of the added transformation is equal to or smaller than D/M , synchronisation can be preserved. Note that the delay elements in the subband domain become of the type complex. In practical SBR embodiments $M=32$. M may also be equal to N .

Please note that in practical embodiments, each transform T comprises two MDCTs and each inverse transform T^{-1} comprises two IMDCTs, as described above.

The lower subbands, in which the transformation T is introduced, are covered by the core decoder. However, although they are not processed by the envelope adjuster of the SBR tool, the high frequency generator of the SBR tool may require their samples in the replication process. Therefore the samples of these lower subbands also need to be available 'non-transformed'. This requires an extra (again complex) delay of DT subband samples in these subbands. The mixing operation performed on the real values and on the complex values of the complex samples may be equal.

Fig. 5 shows an advantageous audio decoder according to an embodiment of the invention, which provides parametric stereo. The bit-stream is split into mono parameters/coefficients and stereo parameters. First a conventional mono decoder is used to obtain the (backwards compatible) mono signal. This signal is analyzed by means of a sub-band filter bank splitting the signal into a number of sub-band signals. The stereo parameters are used to process the sub-band signals to two sets of sub-band signals, one for the left and one for the right channel. Using two sub-band synthesis filters these signals are transformed

PHNL030459EPS

18.04.2003

to the time domain resulting in a stereo (left and right) signal. The stereo processing block is shown in Fig. 3.

Fig. 6 shows an advantageous audio decoder according to an embodiment of the invention, which combines parametric stereo with SBR. The bit-stream is split into mono parameters/coefficients, SBR parameters and stereo parameters. First a conventional mono decoder is used to obtain the (backwards compatible) mono signal. This signal is analyzed by means of a sub-band filter bank splitting the signal into a number of sub-band signals. Using the SBR parameters more HF content is generated, possibly using more sub-bands than the analysis filter bank. The stereo parameters are used to process the sub-band signals to two sets of sub-band signals, one for the left and one for the right channel. Using two sub-band synthesis filters these signals are transformed to the time domain resulting in a stereo (left and right) signal. The stereo processing block is shown in the block diagram of Fig. 3.

It should be noted that the above-mentioned embodiments illustrate rather than limit the invention, and that those skilled in the art will be able to design many alternative embodiments without departing from the scope of the appended claims. In the claims, any reference signs placed between parentheses shall not be construed as limiting the claim. The word 'comprising' does not exclude the presence of other elements or steps than those listed in a claim. The invention can be implemented by means of hardware comprising several distinct elements, and by means of a suitably programmed computer. In a device claim enumerating several means, several of these means can be embodied by one and the same item of hardware. The mere fact that certain measures are recited in mutually different dependent claims does not indicate that a combination of these measures cannot be used to advantage.

CLAIMS:

1. A method of synthesizing an output audio signal based on an input audio signal, the input audio signal comprising a plurality of input subband signals, the method comprising the steps of:
transforming (T) at least one input subband signal from subband domain to
5 frequency domain to obtain at least one respective transformed signal,
delaying ($D_{0...n}$) and transforming the at least one input subband signal to
obtain at least one respective transformed delayed signal;
deriving (P) at least two processed signals from the at least one transformed
signal and the at least one transformed delayed signal,
10 inverse transforming (T^{-1}) the processed signals from frequency domain to
subband domain to obtain respective processed subband signals, and
synthesizing the output audio signal from the processed subband signals.
2. A method as claimed in claim 1, wherein the transforming is a cosine
15 transforming and the inverse transforming is an inverse cosine transforming.
3. A method as claimed in claim 1, wherein the input subband signals comprise
complex samples and wherein a real value of a given complex sample is transformed in a first
transform and a complex value of the given complex sample is transformed in a second
20 transform.
4. A method as claimed in claim 3, wherein the first transform and the second
transform are separate but equal transforms.
- 25 5. A method as claimed in claim 1, wherein the processing comprises a matrixing
operation.
6. A method a claimed in claim 1, wherein the processing comprises a rotation
operation.

7. A method as claimed in claim 1, wherein the at least one subband signal includes the subband signal having the lowest frequency.

5 8. A method as claimed in claim 7, wherein the at least one subband signal consists of 2 to 8 subband signals.

9. A method as claimed in claim 1, wherein the synthesizing step is performed in a subband filter bank for synthesizing a time domain version of the output audio signal from
10 the processed subband signals.

10. A method as claimed in claim 9, wherein the subband filter bank is a complex subband filter bank.

15 11. A method as claimed in claim 9, wherein the complex subband filter bank is a complex Quadrature Mirror Filter bank.

12. A method as claimed in claim 1, wherein the input audio signal is a mono audio signal and the output audio signal is a stereo audio signal.

20

13. A method as claimed in claim 1, the method further comprising:
obtaining a correlation parameter indicative of a desired correlation between a first channel and a second channel of the output audio signal, wherein the processing is arranged for obtaining the processed signals by combining the transformed signal and the
25 transformed delayed signal in dependence on the correlation parameter, and wherein the first channel is derived from a first set of processed signals and the second channel from a second set of processed signals.

14. A method as claimed in claim 13, wherein the processed signals each comprise
30 a plurality of output subband signals, and wherein a first time domain channel and a second time domain channel are synthesized on the basis of the output subband signals respectively, preferably in respective synthesis subband filter banks.

15. A method as claimed in claim 1, wherein the method further comprises:

deriving M subbands to generate M filtered subband signals on the basis of a time domain core audio signal,

generating a high frequency signal component derived from the M filtered subband signals, the high frequency signal component having $N-M$ subband signals, where
5 $N > M$, the $N-M$ subband signals including subband signals with a higher frequency than any of the subbands in the M subbands, the M filtered subbands and the $N-M$ subbands together forming the plurality of input subband signals.

16. A device for synthesizing an output audio signal based on an input audio
10 signal, the input audio signal comprising a plurality of input subband signals, the device comprising:

means for transforming (T) at least one input subband signal from subband domain to frequency domain to obtain at least one respective transformed signal,

means for delaying ($D_{0...n}$) and transforming the at least one input subband
15 signal to obtain at least one respective transformed delayed signal;

means for deriving (P) at least two processed signals from the at least one transformed signal and the at least one transformed delayed signal,

means for inverse transforming (T^{-1}) the processed signals from frequency domain to subband domain to obtain respective processed subband signals, and

20 means for synthesizing the output audio signal from the processed subband signals.

17. An apparatus for supplying an output audio signal, the apparatus comprising:
an input unit for obtaining an encoded audio signal,

25 a decoder for decoding the encoded audio signal to obtain a decoded signal including a plurality of subband signals,

a device as claimed in claim 16 for obtaining the output audio signal based on the decoded signal, and

an output unit for supplying the output audio signal.

30

18. A computer program product including code for instructing a computer to perform the following steps:

transforming (T) at least one input subband signal from subband domain to frequency domain to obtain at least one respective transformed signal,

PHNL030459EPS

10

18.04.2003

delaying ($D_{0...n}$) and transforming the at least one input subband signal to
obtain at least one respective transformed delayed signal;

deriving (P) at least two processed signals from the at least one transformed
signal and the at least one transformed delayed signal,

5

inverse transforming (T^{-1}) the processed signals from frequency domain to
subband domain to obtain respective processed subband signals, and

synthesizing the output audio signal from the processed subband signals.

ABSTRACT:

Synthesizing an output audio signal is provided based on an input audio signal, the input audio signal comprising a plurality of input subband signals, wherein at least one input subband signal is transformed (T) from subband domain to frequency domain to obtain at least one respective transformed signal, wherein the at least one input subband signal is
5 delayed and transformed (D, T) to obtain at least one respective transformed delayed signal, wherein at least two processed signals are derived (P) from the at least one transformed signal and the at least one transformed delayed signal, wherein the processed signals are inverse transformed (T^{-1}) from frequency domain to subband domain to obtain respective
10 processed subband signals, and wherein the output audio signal is synthesized from the processed subband signals.

Fig. 3

PENL030459

1/3

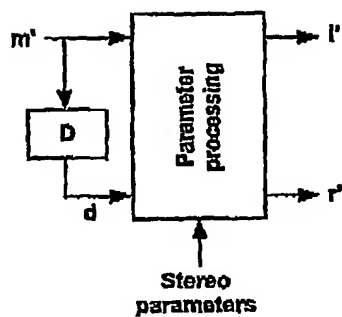


Fig. 1

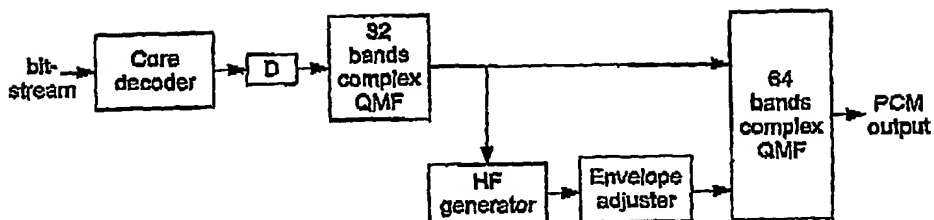


Fig. 2

PHNL030459

2/3

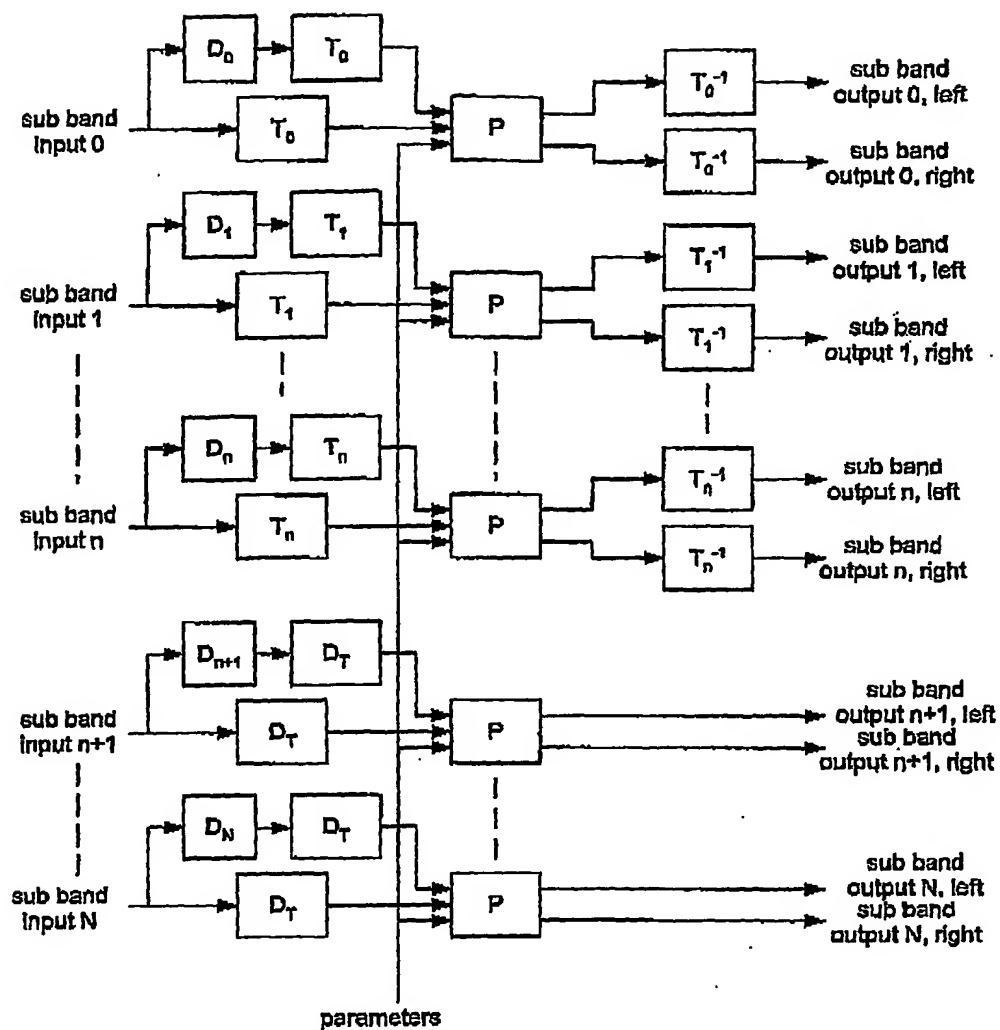


Fig. 3

PHNL030459

3/3

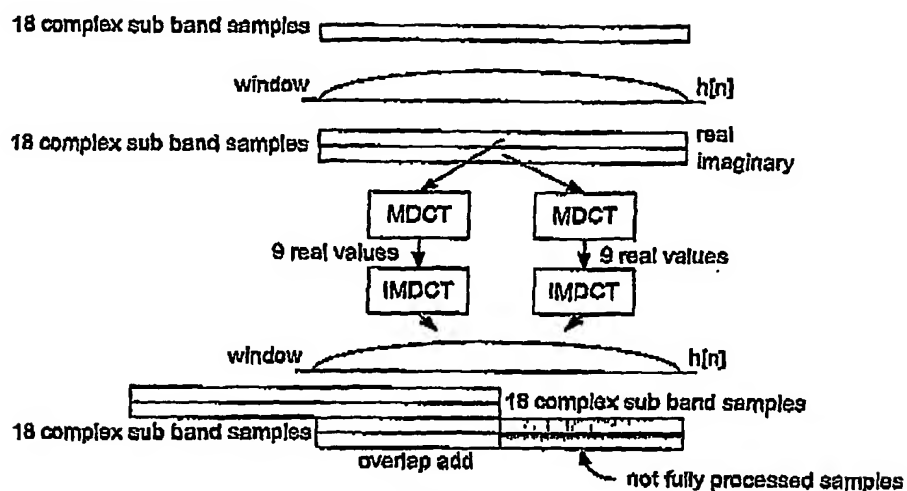


Fig. 4

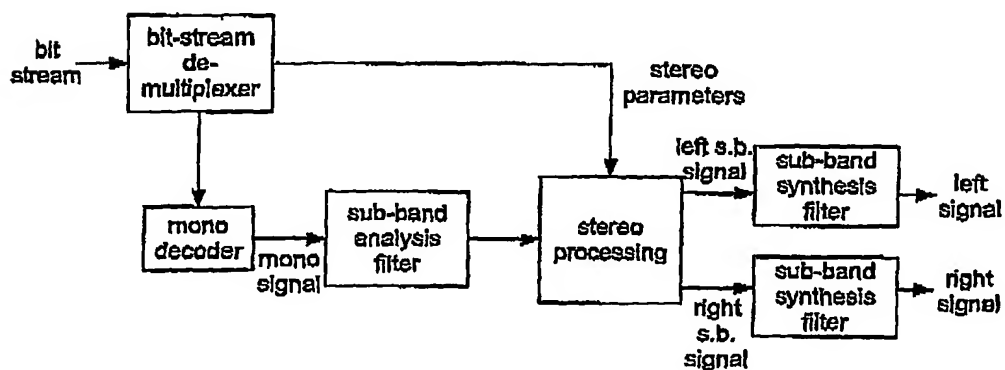


Fig. 5

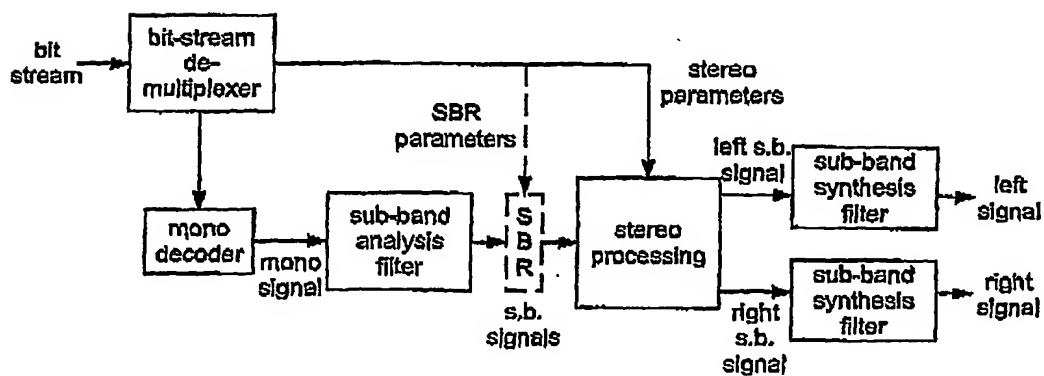


Fig. 6

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☒ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.